Supplementary material

for the paper "TRANSCompel® – a database on composite regulatory elements in eukaryotic genes"

by Olga V. Kel-Margoulis, Alexander E. Kel, Ingmar Reuter, Igor V. Deineko, and Edgar Wingender.

1. CE's containing Smad binding sites.
2. CE's containing SREBP binding sites.
3. CE's containing SF-1 binding sites.
4. Classification of the CEs according to the specific function they provide.

Smad TFs are signal transducers for members of the TGF-β superfamily. Smad TFs are composed of N-terminal Mad-homology domain-1 (DNA-binding domain) and a C-terminal MH2 domain (activation domain).
Smads have been shown to bind to the sequence 5'-GTCG-3'.

CE's containing Smad binding sites.

TRANSCompel acc.*	Cooperating TFs	Functional classification of CE's	Gene
C00194	Smad3/4 / Lef1	Inducible / inducible; cross-coupling of TGF-β and β-cathenin pathways at the early embryonic stages.	twn, *Xenopus laevis*
C00195	Smad3/4 / AP-1	Inducible / inducible	c-jun, *Homo sapiens*
C00196	Smad3/4 / PEBP2	Inducible / inducible & tissue-restricted	Germline Ig C α, *Mus musculus*
C00197	Smad3/4 / CREB	Inducible / inducible	
C00198	Smad3 / TEF3	Inducible / constitutive	PAI 1, *Homo sapiens*
C00199	Smad3/4 / Smad3/4	Inducible / inducible	
C00200	Smad2/4 / FAST-1	Inducible/developmental-stage dependent; FAST is a novel family of fork-head TFs whose expression is restricted to early embriogenesis; activin-responsive element	mix.2, *Xenopus laevis*
C00201	Smad2/4 / FAST-2		goosecoid homeobox gene, *Mus musculus*
C00202	Smad1/4 / OAZ	Inducible/developmental-stage dependent; BMP signaling modulates both factors. OAZ contains 30 zinc fingers.	Xvent-2, *Xenopus laevis*
C00241	Smad2/4 / c-Jun/c-Fos	Inducible / inducible	Smad7, *Mus musculus*
C00242	Smad3/4 / c-Jun/c-Fos	Inducible / inducible	interstitial collagenase, *Homo sapiens*

C00243	Smad2/4 / Sp1	Inducible/constitutive; constitutive factor Sp1 is essential for TGF-β-inducible transcription.	p15INK4B inhibitor of cdks, *Homo sapiens*
C00244	Smad3/4 / Sp1		α2(1) collagen, *Homo sapiens*
C00245	Smad3 / TEF3	Inducible / constitutive	Smad7, *Mus musculus*

* under this accession number one can find detailed information about composite element at http://www.gene-regulation.com/pub/databases.html#transcompel To find all CE's containing Smad binding sites one should search database in the field "Factor" for term "smad".

The ***SREBP*** (Sterol Regulatory Element Binding Protein) family of TFs provides a coordinated transcriptional regulation of genes encoding enzymes of both cholesterol and fatty acid metabolism in response to cellular lipid levels.
Two mammalian genes encoding SREBPs are known: *srebp-1* and *srebp-2*. They are expressed in several tissues, at exceedingly high levels in brown fat. There is a positive autoregulatory loop: transcription of *srebp-1* and *srebp-2* genes is up-regulated by SREBP-1 and SREBP-2 factors.
SREBP-1 is shown to functionally cooperate and physically interact with constitutive transcription factors Sp-1, NF-Y to provide synergistic up-regulation of genes involved in cholesterol and fatty acids biosynthesis.
SREBPs belong to the bHLH-ZIP superclass of TFs and are known to bind the DNA motif 5'-ATCACCCCAC-3'.

CE's containing SREBP binding sites.

TRANSCompel acc.*	**Cooperating TFs**	**Functional classification of CE's**	**Gene**
C00204	SREBP-1 / Sp-1	Inducible / constitutive	LDL receptor, *Homo sapiens*. Product of this gene is essential for cholesterol uptake.
C00205			Fatty acid synthase, *Rattus norvegicus*. This gene encodes a multi-functional enzyme of fatty acid biosynthesis.
C00206	SREBP-1 / NF-Y	Inducible / constitutive	Fatty acid synthase, *Rattus norvegicus*
C00207			Farnesyl diphosphate synthase, *Rattus norvegicus*
C00208			3-hydroxy-3-methylglutaryl-CoA synthase, *Mesocricetus auratus*. This gene encodes a rate-limiting enzyme of cholesterol biosynthesis.
C00210			srebp-2, *Homo sapiens*
C00211			Glycerol-3-phosphate acyltransferase, *Mus musculus*
C00209	SREBP-1 / CREB	Inducible / inducible	3-hydroxy-3-methylglutaryl-CoA synthase, *Mesocricetus auratus*

* under this accession number one can find detailed information about composite element at http://www.gene-regulation.com/pub/databases.html#transcompel To find all CE's containing SREBP binding sites one should search database in the field "Factor" for term "srebp".

SF-1 (Steroidogenic factor 1) is an orphan nuclear receptor, whose expression is restricted to the steroidogenic cells of the adrenal gland and gonads and to the pituitary gonadotropes. This transcription factor is known to play an important role in reproductive development and function.
SF-1 binds as a monomer to its cognate DNA binding site 5'-TGACCTTG-3'.

CEs containing SF-1 binding sites.

TRANSCompel acc.*	**Cooperating TFs**	**Functional classification of CE's**	**Gene**
C00213, C00214	SF-1 / Egr-1	Tissue-restricted / inducible; Egr-1 is transiently activated by a variety of external stimuli such as serum or growth factors.	Luteinizing hormone β subunit, *Rattus norvegicus*
C00215	SF-1 / Ptx1	Tissue-restricted / Tissue-restricted. Pan-pituitary factor, Ptx1, collaborates with a lineage-restricted factor, SF-1, for activation a cell-specific program of gene expression	Luteinizing hormone β subunit, *Bos taurus*
C00217	SF-1 / ER	Tissue-restricted / inducible; cell type-restricted induction by oestrogenes.	Gonadotropin II βsubunit, *Oncorhynchus tschawytscha*
C00218	SF-1 / Sp1	Tissue-restricted / ubiquitous. Ubiquitous factor Sp1 is essential for tissue-restricted gene expression.	cyp11A, *Bos taurus*
C00219	SF-1 / SOX-9	Tissue-restricted / tissue- & developmental-stage restricted; SOX-9 expression is restricted to the embryonic Sertoli cells.	Anti-Mullerian hormone *(*Mullerian inhibiting substance*)*, *Homo sapiens*
C00220	SF-1 / GATA-4	Tissue-restricted / tissue- & developmental-stage restricted; GATA-4 is early marker of the embryonic Sertoli cells.	Anti-Mullerian hormone *(*Mullerian inhibiting substance*), Mus musculus*
C00221	SF-1 / Sp1	Tissue-restricted / ubiquitous. Ubiquitous factor Sp1 is essential for tissue-restricted gene expression.	Steroidogenic acute regulatory gene, *Homo sapiens*
C00222	SF-1 / CREB	Tissue-restricted / inducible;	Inhibin-α gene*, Rattus norvegicus*
C00223	SF-1 / Egr-1	Tissue-restricted / inducible; Egr-1 is transiently activated by a variety of external stimuli such as serum or growth factors.	Luteinizing hormone β subunit, *eguine*

C00225	SF-1 / CREB	Tissue-restricted / inducible. This CE provides high level of aromatase gene expression in Leydig cells upon cAMP induction.	cyp19 (P450 aromatase), *Rattus norvegicus*

* under this accession number one can find detailed information about composite element at http://www.gene-regulation.com/pub/databases.html#transcompel To find all CE's containing SF-1 binding sites one should search database in the field "Factor" for term "SF-1".

Classification of the CE's according to their specific function (illustrative examples).

Starting with the TRANSCompel 4.4 functional classification of the composite elements is one of the important features of the database. One can search database in the field "Functional classification" for the following terms corresponding to the first level of classification: inducible/inducible, inducible/tissue-restricted, inducible/constitutive, tissue-restricted/ubiquitous, tissue-restricted/tissue-restricted, constitutive/ constitutive; as well as for the terms corresponding to the second level of classification: T-cells, B-cells, myeloid, pituitary, liver, placenta, muscle, immune and some others.

Functional properties TF1	Functional properties TF2	Gene	Location of CE	TRANSCompel acc.*
		1) "inducible/inducible"		
c-Ets, Ras-dependent ind.	AP-1, ind. PKC	SR; macrophage scavenger receptor gene, *Hs*	-65 … -52	C00079
C/EBPβ, ind. IL-6	NF-κB, ind. IL-1 and TNFα	serum amyloid A2, *Hs*	-179 … -82	C00100
NF-AT, ind. Ca^{2+}	AP-1, ind. PKC	interleukin-2, *Hs*	-287 … -266	C00109
C/EBPβ, ind. IL-6	AP-1, ind. PKC	TNFα, *Hs*	-107 … -74	C00178
IRF-1, ind. by interferon α и γ	NF-κB, ind. IL-1 and TNFα	interferon β, *Hs*	-77 … -55	C00061
		2) "inducible/tissue-restricted"		
HNF-1, hepatocyt.	C/EBPβ, ind. IL-6	β-fibrinogen gene, *Hs*	-133 … -77	C00095
HNF-3, hepatocyt.	GR, ind. Glucocort.	tyrosine amino transferase, *Rn*	-2509... -2430	C00128
HNF-4, hepatocyt.	CREB, ind. cAMP		-3650… -3586	C00129
AML1, T- and myeloid cells	c-Ets, Ras-dependent ind.	T-cell receptor β, *Hs*	*3′ энхансер*	C00020

Pit-1, pituitary	c-Ets, Ras-dependent ind.	prolactin, *Rn,*	-162 … -147 -217… -190	C00137 C00131
3) "tissue-restricted/ubiquitous"				
Myogenin, muscle cells	Sp1, ubiquitous	acetylcholine receptor α-subunit, *Hs*	-89 … -47	C00027
Pit-1, pituitary	Sp1, ubiquitous	growth hormone gene, *Hs*	-139 … -105	C00038
HNF-1, hepatocyt.	Oct-1, ubiquitous	large surface antigen, *HBV*	-86 … -51	C00048
C/EBP α, hepatocytes	NF-Y, ubiquitous	serum albumin, *Mm*	-110 … --80	C00069
4) "inducible/constitutive"				
NF-κB, ind. IL-1 and TNFα	Sp1, constitutive	HIV-1 LTR	-90 … -68	C00055
c-Ets, ind. by Ras	Sp1, constitutive		-140 … -127	C00007
C/EBPβ, ind. IL-6	Sp1, constitutive	CYP2D5, *Rn*	-105 … -83	C00070
Stat 3, ind. IL-6	Sp1, constitutive	transcription factor C/EBPδ, *Mm*	-120 … -102	C00179
GR, ind. glucocort.	Oct-1, constitutive	MMTV LTR	-89 … -49	C00043
5) "tissue-restricted/tissue-restricted"				
HNF-4 - liver, gut, kidney	C/EBPα - liver, gut, adipocytes, brain, myelocytes	apolipoprotein B, *Hs*	-81 … -52	C00122
C/EBPα - liver, gut, adipocytes, brain, myelocytes	AML1, T- and myeloid cells	M-CSF receptor (*c-fms*), *Hs*	-84 … -67	C00145
MEF2A, muscle cells	Myogenin, muscle cells	muscle-restricted transcription factor MRF4, *Rn*	-26 … +27	C00120

* under this accession number one can find detailed information about composite element at http://www.gene-regulation.com/pub/databases.html#transcompel